View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Waterloo's Institutional Repository

Asymptotic Distribution of the Optimal Value in Random Linear Programs: Application to Maximum Expected Shortfall

by

Jesse Hall

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Quantitative Finance

Waterloo, Ontario, Canada, 2020

© Jesse Hall 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The properties of risk measures are of fundamental concern in quantitative finance, particularly in times of uncertainty. We study the behaviour of the asymptotic distribution of the maximum expected shortfall of a portfolio that has both market and credit risk, where the marginal distributions of the risk factors are known but their joint distribution is unknown. We study the limiting behaviour of linear programs, as the maximum expected shortfall has a form similar to an optimal transport problem with stochastic cost function, and derive a result for the asymptotic distribution of the optimal solution similar to the central limit theorem. We then present simulations of maximum expected shortfall for a portfolio consisting of two counterparties subject to credit and market risk using the Basel IRB approach and a Merton single-factor copula model for portfolio losses. We observe that the histogram of maximum expected shortfall is well-described by a generalized extreme value distribution with a negative shape parameter.

Acknowledgements

I would like to thank my wife Sadie, for being supportive and caring throughout this long and extended process. My kids, James and Olivia, for understanding as best they can why Daddy has to work all day and then again all night. My mum has been amazing, coming down to look after the kids and take care of my family while I've been either writing this thesis or deployed by the army. My dad is a fountain of advice and encouragement without whose influence I likely never would have pursued education as far as I have.

The University of Waterloo, and the MQF program in particular, have been extremely accommodating and supportive throughout this process. My supervisor, David Saunders, has shepherded me through this process far more than I could ever have asked, and his encouragement and steadfast support has frankly been the only thing that has seen me through to the end. The professors who have taught me and been incredibly forebearing and understanding as I have struggled with balancing my family, my other commitments, and my health. The support staff, Mary-Lou and Helen, who have made sure the i's get dotted and the t's crossed, whose professionalism and dedication to their duties allow them to bear the weight of we students upon their shoulders.

Dedication

For my children.

Table of Contents

List of Figures

Chapter 1

Introduction

The measurement and quantification of risk are topics of great importance to the financial industry, and any progress towards greater understanding of this quantity is potentially highly valuable. Risk management is hardly an exact science, and requires both a thorough understanding of what exactly is being measured and how it is being modelled. The stakes are enormous - failures of risk management were identified [7] as a key contributing factor to the worldwide financial crisis of 2008.

Risk measures that in some way quantify losses of a portfolio are an essential element of any portfolio strategy. Among the most popular is Value at Risk (VaR):

$$VaR_\alpha(X) = min\{z|F_X(z) \geq \alpha\} \tag{1.1}$$

where X is a loss random variable and $\alpha \; \epsilon \;]0,1[$.

The use of VaR as a standard risk measure can trace its origins to an alignment of factors. VaR is fairly straightforward - it is simply the minimum loss at a particular threshold probability, and it can be expressed in plain language in a relatively comprehensible way - a 1% VaR of $1,000,000 means there is a 1% chance of losing at least a million dollars or, equivalently, there is a 99% chance of losing no more than a million dollars. It is easy to calculate; one simply simulates expected returns for a portfolio, orders the returns from best to worst, and picks a threshold probability. The VaR of the portfolio is then the return of the first scenario beyond the threshold. So, in a set of 1,000 simulations of returns ordered from best to worst, the 99% VaR is the return of the 990th simulation.

The ease of understanding and calculation led to VaR dominating as a risk measure for and becoming enshrined as the preferred risk measure for market risk when determining capital requirements in the Basel II accord. VaR is not without its drawbacks, however, and these are not insignificant. It is not a coherent risk measure - it is generally not sub-additive, so the VaR of a portfolio can be greater than the sum of the VaRs of its constituent parts. Moreover, VaR is difficult to optimize, especially in discrete rather than continuous cases, and outside the assumption of normally-distributed losses even in the case of continuous loss functions [20].

Expected shortfall (ES), also referred to as conditional value-at-risk (CVaR), measures the expected loss given that loss exceeds VaR:

$$ES = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_{\beta}\ d\beta \tag{1.2}$$

It has a number of theoretical advantages over VaR - it is spectral, convex, and coherent, and it can be solved practically as a linear programming problem [20]. Moreover, it quantifies the loss in a way that VaR does not - VaR is the minimum loss at a given probability, but it says nothing about how much is likely to actually be lost if the loss occurs. ES gives the average loss expected in the scenarios beyond the cutoff probability, thereby offering some idea of how much will actually be lost rather than providing a floor. Because of these factors, ES is being incorporated into the new Basel standards [4] for market risk capital requirements.

1.1 Problem Statement

The focus of this thesis will be the situation in which expected shortfall must be calculated for a loss function $L(X,Y)$ when the marginal distributions of X and Y are known but their joint distribution is unknown. In particular, we study the asymptotic distribution of the maximum expected shortfall when the marginal distributions of X and Y are simulated by observed random variables. Previous work [18] focused on the problem of determining the copula giving the worst-case joint distribution of credit and market factors given known marginals; that is, the joint distribution that results in the greatest ES. The maximum ES can be obtained by solving an optimal-transport type linear program if the marginal distributions of X and Y are known and both have fi

nite range. The question then arises whether the distribution of the maximum ES estimated based on simulated data can be characterized in some meaningful way.

This thesis is broken down into three chapters after the introduction. The first chapter describes the theoretical background of the problem, as well as giving an overview of previous work on the topic. The next chapter describes the method of solving the problem, describes the simulations, and discusses the results. The final chapter summarizes the conclusions and offers some thoughts on directions of future work. The matlab code that generated the results is included as an appendix.

Chapter 2

Theoretical Background

This chapter describes an overview of the existing framework underlying the present work and describes the theoretical contribution of this thesis. We will give an overview of the present state of the literature and some of the key references on the use of expected shortfall in counterparty credit risk evaluations, linear programming and optimal transport problems, limiting distributions for linear programs, and the worst-case expected shortfall. I will then describe the regulatory framework and the calculation of regulatory capital requirements using expected shortfall. This is followed by a description of linear programming and the optimal transport problem in the context of solving the worst-case ES. I will provide an overview of the delta method and Hadamard directional differentiability upon which the present investigation is predicated and by which it was motivated. Finally, I will prove the limiting distribution of the maximum expected shortfall in the case of finite sample spaces.

2.1 Literature Review

The problem of assessing counterparty credit risk in addition to market risk has been given increased prominence in the years following the financial crisis of 2008. Measures of risk that account for both market risk and credit risk are valuable as they allow risk managers to quantify potential losses from different causes under a variety of circumstances. Expected shortfall has the advantage of being coherent and spectral [3, 1]. The situation of two independent sources of risk (credit factors and market factors, for example) for which the marginal distributions of the loss function are known but the joint distribution is unknown has been described in the context of a stress-testing approach [10, 22].

Previous work on the maximum ES, of which this work is a continuation, was reported in Memartoluie *et al.* [18]. This work focused on determining the worst-case ES of a portfolio with known marginal distributions by solving an optimal-transport-linear-like linear program to determine the joint distribution for credit factors and market factors consistent with the marginals that produced the worst expected shortfall for a given (stochastic) loss function. The purpose of the current work, indeed, is to extend this analysis by looking at the limiting behaviour of this maximum ES.

The optimal transport problem is the subject of a large literature in optimization, statistics, and computer science (see [19] and [31] for a thorough discussion). It has also been used for deriving bounds on prices of financial instruments [5]. Applications to risk measures are discussed in chapter 9 of Rüschendorf's book [24] as well as in McNeil *et al.* [17] for specific loss functions. Glasserman and Yang [11] use an optimal transport approach to examine bounds on credit valuation adjustment (CVA).

The limiting behaviour of the solutions of linear programming problems is a subject under rapid development, with previous work on the topic focusing on specific problems or particular cases. Smith [27] studied a linear regression model of the form $Y_i = x_i^T\beta + W_i$, where x_i is a covariate matrix and the W_i's are i.i.d. random variables, and obtained an estimate for β by solving a linear program under some fairly restrictive assumptions about the W_i's. He then derived the limiting distribution for the solutions of the linear program (namely, the $\hat{\beta}_n$'s). This work was expanded by Knight [15], who was able to repeat the derivation without the restrictive assumptions on the W_i's and x_i's that Smith used, showing that the solution to the linear program considered by Smith converges in distribution to the generalized extreme value distribution.

The limiting behaviour of the optimal transport problem has been examined in detail under a number of specific circumstances. Sommerfeld and Munk [28] investigated the optimal transport problem with a stochastic cost function and derived a central-limit-like result for the limiting distribution in the context of statistical inference using a Hadamard directional derivative approach similar to the approach used here. These ideas were further developed in the case of empirical optimal transport problems on finite metric spaces [29, 13]. A central-limit type result was obtained in the case of a quadratic cost function by del Barrio and Loubes [9] and for Wasserstein distances in one dimension [8].

After the work described herein was mostly completed, a preprint [14] appeared describing further work on limit laws for stochastic linear programs. In particular, they prove that the limiting distribution for the stochastic optimal transport problem should be a Gaussian if the dual solution is nondegenerate. In the case of degeneracy in the solutions, however, the limiting distribution can take on a more complex form. The work presented herein is consistent with this finding, insofar as we find the limiting distribution can be well fit with a Gaussian in the simple case of coefficients equal to the sum of two normal distributions but a generalized extreme value distribution in the more complex case of a two-counterparty portfolio.

2.2 Regulatory framework and the basel capital charge

Any time a transaction is entered into that is not settled immediately, it creates exposure between counterparties, and each assumes the risk that the other will default upon their obligation. The Basel Accord (see reference [4] and related documentation, available at bis.org) imposes an obligation to maintain capital reserves against the possibility of counterparty default, and includes standards for using an internal ratings-based method for calculating counterparty credit risk and capital requirements. The Basel capital charge is given by the formula

$$C = \max(EAD - CVA, 0) \cdot LGD \cdot \left[\Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho} \cdot 0.999}{\sqrt{1-\rho}}\right) - PD\right] \cdot MA \tag{2.1}$$

where C is capital, EAD is exposure at default, LGD is loss given default, PD is probability of default, and MA is the maturity adjustment, and CVA is the credit valuation adjustment. The maturity adjustment can be subsumed into the CVA and omitted from the equation where this is the case. The portfolio losses L due to the counterparty's default are then equal to

$$L = EAD \cdot LGD \cdot \mathbb{I}(\text{default}) \tag{2.2}$$

Here $\mathbb{I}$ is the indicator function, which is equal to 1 if the counterparty defaults and 0 otherwise.

In order to calculate ES, we require an expression that can be evaluated for portfolio losses. In the case at hand where losses may be due either to adverse market movements or to a counterparty experiencing a credit event such as a bankruptcy, we encounter the problem of *wrong way risk* (WWR), which occurs when an event that increases exposure to a counterparty (i.e. increases how much money they owe to us) also increases the likelihood of that counterparty experiencing a credit event. Computing losses under different market and credit scenarios must account for these factors.

In order to assess portfolio performance, each counterparty needs to be assigned a *probability of default* (PD). This is not directly measurable for an individual counterparty, but one could use historical default rates for counterparties with similar characteristics to arrive at an estimate.

We define the *creditworthiness index* (CW) of a counterparty so that the counterparty defaults if $CW \leq \Phi^{-1}(PD)$, where Φ^{-1} is the inverse CDF of the standard normal. To model CW, we follow Andersen and Sidenius [2] (who were, in turn, following Vasicek [30]) and use a single-factor model, breaking the sources of change to the CW of a counterparty down into two types: a systematic factor Z that affect all counterparties equally, and idiosyncratic factors ϵ that are specific to the individual counterparty. For simplicity, we assume these are independent standard normal random variables and use a single factor Gaussian copula:

$$CW_i = \sqrt{\rho_i} \cdot Z + \sqrt{1-\rho_i}\, \epsilon_i. \tag{2.3}$$

The parameter ρ is called the *factor loading* of the CW_i, which is the creditworthiness index of the ith counterparty.

This characterizes the portfolio losses due to default. The exposure at default to the counterparty is characterized by a separate random variable Y, which we can take to be independent of the idiosyncratic factors (any wrong-way risk being subsumed into the co-dependence of Y the systematic factor Z).

The conditional expectation of the counterparty's portfolio loss due to the counterparty's default given Z under the market scenario $Y = y_m$ is given by

$$L_m(Z) = y_m \cdot \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho} \cdot Z}{\sqrt{1-\rho}}\right) \tag{2.4}$$

and for multiple counterparties the total loss is the sum of the individual losses. Equipped with this expression, ES can be calculated using the expression [18]:

$$ES_\alpha(L) = \sup_{G \in \mathbb{G}} E_G[L] \tag{2.5}$$

where L is a random variable defined on a probability space with a probability measure P, and $\mathbb{G}$ is the set of all probability measures $G \ll P$ whose density is bounded by $1/(1-\alpha)$.

Thus equipped with an expression for the losses and an expression for ES based on losses, we can look at how to maximize the expression.

2.3 Linear programming and the optimal transport problem

Linear programming aims to solve optimization problems that take the form of a linear objective function (the function to be optimized) with constraints that are linear equalities or inequalities [16]. As it turns out, a great many optimization problems across a wide variety of different fields can be stated in this form or one similar to it, and efficient algorithms exist for solving them. The problem can be written

$$\begin{aligned} \min_x \; c^T x \\ \text{subject to } Ax = b, x \geq 0 \end{aligned} \tag{2.6}$$

where c is the coefficient vector and A is the $m \times n$ constraint matrix, b is an m-vector, x is an n-vector. This is said to be the standard form - all inequality constraints are simply non-negativity constraints, and all remaining constraints are strict equality constraints. If the constraints are consistent, a vector x that satisfies this set of linear equations is called a feasible solution; if $n > m$, there can be multiple feasible solutions.

A linear program in this form, with A of full rank, can be written in terms of an $m \times m$ matrix composed of linearly independent columns of A, which forms a basis. The solution to the set of linear equations, with the other $n - m$ values of x set to zero, is then referred to as a basic solution. The fundamental theorem of linear programming then states that i) if there is a feasible solution, there is a basic feasible solution; and ii) if there is an optimal feasible solution, there is an optimal basic feasible solution [16]. Because the sub-matrix of A forms a basis, this can be interpreted as stating that if a solution exists, one must exist at the corners of a convex polytope consisting of all n-vectors x that satisfy $Ax = b, \; x \geq 0$. If one or more of the values of the basic variables in x are equal to zero, the solution is said to be degenerate.

Every linear program in this form has a dual program, in which the minimization problem is swapped out for a maximization problem. The linear program with inequality constraints in primal form

$$\begin{aligned} \min_x c^T x \\ \text{subject to } Ax \geq b, x \geq 0 \end{aligned} \tag{2.7}$$

has a dual form that can be expressed as

$$\begin{aligned} \max_{y} \ & y^T b \\ \text{subject to } & y^T A \leq c^T, y \geq 0 \end{aligned} \tag{2.8}$$

This is the symmetric expression of the primal and dual; when the primal problem is instead stated in standard form, the dual problem becomes

$$\begin{aligned} \max_{u,v} \ & u^T b - v^T b \\ \text{subject to } & u^T A - v^T A \leq c^T, \\ & u \geq 0, v \geq 0 \end{aligned} \tag{2.9}$$

The strong duality theorem of linear programming states that if the primal problem has a finite optimal solution, then the dual problem does as well; moreover, the values of the objective functions will be the same for the two corresponding solutions.

A particular application of linear programming exists for what is knows as the Optimal Transport Problem (OTP). The essence of the problem is based on the idea of moving piles of earth of size $\mathbf{p} = [p_1, p_2, ..., p_m]$ from from a set of m starting locations to make piles of size $\mathbf{q} = [q_1, q_2, ..., q_n]$ at n destination locations. The object is to minimize a cost function $\mathbf{c} \in \mathbb{R}^{m \times n}$ that quantifies the cost of moving dirt from a starting point p_i to a destination point q_j. If π_{ij} is the amount of earth moved from p_i to q_j, the linear program becomes:

$$\min_{\pi} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij} \tag{2.10}$$

$$\sum_{j=1}^{n} \pi_{ij} = p_i, \ i \in \{1, \ldots, m\} \tag{2.11}$$

$$\sum_{i=1}^{m} \pi_{ij} = q_j, \ j \in \{1, \ldots, n\} \tag{2.12}$$

$$\pi_{ij} \geq 0, \ \forall \ i, j \tag{2.13}$$

The constraints guarantee that the total amount of dirt (or whatever it is) delivered to the destinations is equal to the total amount removed from the starting locations, that all of the dirt is moved from the starting locations, and that each destination receives exactly the correct amount. This yields a linear program with mn terms in the objective function and $m + n + mn$ constraints.

The dual of this problem can be thought of as trying to optimize the cost function rather than the transportation plan. It can be stated as

$$\max_{\phi,\psi} \sum_{i=1}^{m} \phi_i p_i + \sum_{j=1}^{n} \psi_j q_j \tag{2.14}$$

$$\phi_i + \psi_j \leq c_{ij}, \ \forall \ i, j \tag{2.15}$$

since the primal problem was stated in its standard form (i.e. only equality constraints, strict non-negativity).

2.4 The limit theorem for maximum expected shortfall

We take as our starting point the multivariate central limit theorem. Throughout this section, $N_k(\mu, \Sigma)$ will be the k-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ, $\Phi(x)$ will denote the pdf of the standard normal distribution evaluated at x, and "$\rightsquigarrow$" indicates convergence in distribution.

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random vectors in $\mathbb{R}^k$ with mean vector $\mu = \mathbb{E}[Y_1]$ and covariance matrix $\Sigma = \mathbb{E}[(Y_1 - \mu)(Y_1 - \mu)^T]$. Then the central limit theorem states the following:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu) = \sqrt{n}(\bar{Y}_n - \mu) \rightsquigarrow N_k(0, \Sigma) \tag{2.16}$$

Suppose we have two finite sets with cardinality M_p and M_q, and probability vectors $p \in \mathbb{R}^{M_p}, q \in \mathbb{R}^{M_q}$. Let a random vector ξ be equal to the ith basis vector in $\mathbb{R}^{M_p}$ with probability p_i, and similarly let η be equal to the jth basis vector in $\mathbb{R}^{M_q}$ with probability q_j, where ξ and η are independent. The random probability vector $\hat{p}^n = \bar{\xi}_n$ (or $\hat{q}^n = \bar{\eta}^q_n$) is called the empirical measure of the sample. The multivariate central limit theorem leads to the following result:

$$\sqrt{n}\left(\begin{bmatrix}\hat{p}^n\\ \hat{q}^n\end{bmatrix}-\begin{bmatrix}p\\ q\end{bmatrix}\right)\rightsquigarrow N_M(0,\Sigma) \tag{2.17}$$

with $M = M_p + M_q$. Here the covariance matrix Σ is determined by $\mathrm{var}(\xi_i) = p_i(1-p_i)$ and $\mathrm{cov}(\xi_i, \xi_k) = -p_i p_k$ if $i \neq k$, with similar expressions for η, and the fact that ξ and η are independent.

We now consider this in the context of a financial scenario. Suppose we have two risk factors, Y and Z, representing market risk factors and credit risk factors respectively, whose marginal distributions are known but whose joint distribution is unknown. The total number of possible values for Y is m, with probabilities given by the vector p, and the total number of possible values for Z is n, with probabilities given by q. The loss given that Y takes its i^{th} value y_i and Z takes its j^{th} value z_j is L_{ij}. Consider a given joint distribution π, with π_{ij} giving the probability that $Y = y_i$ and $Z = z_j$. The expected shortfall of L under π is then (see [18]):

$$\mathrm{ES}_\alpha(L) = \max_{\Theta\in\Theta^\alpha}\sum_{i=1}^{m}\sum_{j=1}^{n} L_{ij}\cdot\Theta_{ij} \tag{2.18}$$

where Θ^α is the set of all probability measures whose density $\frac{d\Theta}{d\pi}$ is bounded by $\frac{1}{1-\alpha}$. The problem of maximizing the expected shortfall among all joint distributions π with marginal distributions p and q then becomes:

$$\max_{\pi,\Theta}\sum_{i=1}^{m}\sum_{j=1}^{n}(L_{ij}\Theta_{ij} + 0\cdot\pi_{ij}) \tag{2.19}$$

$$\sum_{j=1}^{n}\pi_{ij} = p_i,\ i\in\{1,\dots,m\} \tag{2.20}$$

$$\sum_{i=1}^{m}\pi_{ij} = q_j,\ j\in\{1,\dots,n\} \tag{2.21}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\Theta_{ij} = 1 \tag{2.22}$$

$$\Theta_{ij} \leq \frac{\pi_{ij}}{(1-\alpha)},\ i\in\{1,\dots,m\}, j\in\{1,\dots,n\} \tag{2.23}$$

$$\pi_{ij},\Theta_{ij}\geq 0,\ i\in\{1,\dots,m\}, j\in\{1,\dots,m\} \tag{2.24}$$

The constraints come from the requirement that, by definition, the total probability of all scenarios contributing to α-level ES must be $1 - \alpha$ (i.e. Θ is a probability measure).

This puts the worst-case joint distribution in the form of an optimal-transport-like problem, which can be solved as a linear program using an appropriate commercial solver software. The problem has $2nm$ terms in the objective function, and $n + m + nm + 1$ constraints (not including the non-negativity constraints).

The dual problem takes the form

$$\min_{\phi,\psi,\rho,\beta} \sum_{i=1}^{m} \phi_i p_i \sum_{j=1}^{n} \psi_j q_j + \beta \tag{2.25}$$

$$(1-\alpha)(\phi_i - \psi_j) - \rho_{ij} \geq 0,\ i \in \{1, \dots, m\}, j \in \{1, \dots, n\} \tag{2.26}$$

$$\rho_{ij} + \beta \geq L_{ij},\ i \in \{1, \dots, m\}, j \in \{1, \dots, n\} \tag{2.27}$$

$$\rho_{ij} \geq 0,\ i \in \{1, \dots, m\}, j \in \{1, \dots, n\} \tag{2.28}$$

The primal problem has a bounded non-empty feasible set, so the dual also has a non-empty feasible solution ($\pi_{ij} = p_i q_j,\ \Theta = \pi$ is feasible) and their optimal values are equal by the duality theorem discussed above. If $(\phi^k, \psi^k, \rho^k, \beta^k)$ are the vertices of the dual polytope where we fix $\phi_1^k = 0$, then there are finitely many possible solutions. The optimal value of the dual objective function becomes

$$M_D(p, q) = \min_k (\phi^k \cdot p + \psi^k \cdot q + \beta^k) \tag{2.29}$$

It is this quantity for which we wish to find a limiting distribution. This will be done as follows: the Delta method [21] gives a limit theorem for $g(Y_n)$ by essentially employing a first-order Taylor expansion to a limit theorem for Y_n. Given a sequence $\{t_n | n \in \mathbb{N}\}$ such that $t_n(Y_n - \mu) \rightsquigarrow Z$, then

$$t_n(g(Y_n) - g(\mu)) = t_n \nabla g(\mu)(Y_n - \mu) + t_n R(Y_n) \tag{2.30}$$

$$\rightsquigarrow \nabla g(\mu) \cdot Z. \tag{2.31}$$

When Z is normal, then so is $\nabla g(\mu) \cdot Z$.

In this context, for the Delta method to apply, the correct notion of differentiability is Hadamard directional differentiability. We can begin by defining the Gateaux derivative [6]:

Definition 1. *Given a convex set $C \subseteq \mathbb{R}^M$ and a function $g : C \to \mathbb{R}$, the Gateaux directional derivative of g at a point $\bar{x}$ in the direction $d \in \mathbb{R}^M$ is:*

$$g'(x; d) = \lim_{t \searrow 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \tag{2.32}$$

This does not guarantee, however, that given Gateaux differentiable functions $g_1(x), ..., g_n(x)$ the max function $g(x) = \max_{i \in \{1,...,n\}}\{g_i(x)\}$ will also be Gateaux differentiable. Fortunately, we can make use of the following [6]:

Proposition 1. *For points $\bar{x}$ in the interior of $C \subset \mathbb{R}^N$, if we define an index set $K = \{i | g_i(\bar{x}) = g(\bar{x})\}$ we have the following result for the Gateaux directional derivative of the max function $g'(\bar{x})$ in the direction d:*

$$g'(\bar{x}; d) = \max_{i \in K}\{\langle \nabla g_i(\bar{x}), d \rangle\} \tag{2.33}$$

As the maximum ES problem is a maximization (one can swap max and min by inverting the sign of the coefficient matrix if need be), a version of this is required to ensure the differentiability of the maximum ES problem as the parameter vectors p and q are perturbed. The Gateaux derivative of the function M_D from 2.29, which gives the optimal value of the when the primal problem is feasible in the direction (d_p, d_q) is:

$$M'_D(p, q; d_p, d_q) = \min_{n \in I(p,q)} (\phi^n \cdot d_p + \psi^n \cdot d_q) \tag{2.34}$$

where $I(p, q)$ is the set of all values which minimize $M_D(p, q)$.

In order to obtain a limit theorem for $g(Y_n)$, we need g to satisfy the stronger condition of Hadamard directional differentiability [23]:

Definition 2. *Let $C \subseteq \mathbb{R}^K$. The function $g : \mathbb{R}^K \to \mathbb{R}$ is said to be Hadamard directionally differentiable at $x \in \mathbb{R}^K$ tangentially to C if the limit*

$$g'(\bar{x}; d) = \lim_{n \to \infty} \frac{g(\bar{x} + t_n d_n) - g(\bar{x})}{t_n} \tag{2.35}$$

exists for all sequences $\{d_n\}$ converging to d of the form $d_n = t_n^{-1}(x_n - x)$, where $x_n \in C$ and $t_n \searrow 0$.

If the Hadamard derivative exists for all sequences $\{d_n | n \in \mathbb{N}\}$ then $g(x)$ is Hadamard differentiable at $\bar{x}$. Furthermore, we have the following [26]:

Proposition 2. *If C is a convex set and a locally Lipschitz mapping g is Gateaux directionally differentiable at $\bar{x} \in C$ tangentially to C, then g is also Hadamard directionally differentiable at $\bar{x}$ and the two derivatives are identical.*

In the case under consideration, it is sufficient that the function g be Hadamard directionally differentiable tangentially to C - for the optimal transport problem and the maximum ES problem, we need only consider directions that correspond to feasible problems - p and q must still be probability vectors. Since $x + t_n d_n = x_n \in C$ for all such sequences of d_n, this constitutes a feasible problem.

In the case where we have a function g that is Hadamard directionally differentiable, we have the following theorem [21]:

Theorem 1 (The Delta Method). *Given $\bar{x} \in \mathbb{R}^K$, $C \subseteq \mathbb{R}^K$, a sequence of random vectors $\{\xi_n\}_{n\in\mathbb{N}}$ such that for n large enough $\xi_n \in C$ almost surely, and a real sequence $\{a_n\}_{n\in\mathbb{N}}$ with $a_n \to \infty$ such that $a_n(\xi_n - \bar{x}) \rightsquigarrow V$ for a random vector V, and a mapping $g : C \to \mathbb{R}$ that is Hadamard directionally differentiable at $\bar{x}$ tangentially to C, then*

$$a_n(g(\xi_n) - g(\bar{x})) \rightsquigarrow g'(\bar{x}; V). \tag{2.36}$$

The strategy is to derive a limit theorem by applying the above result with

$$g = M_D, \quad \bar{x} = \begin{pmatrix} p \\ q \end{pmatrix}, \quad \xi_n = \begin{pmatrix} p^n \\ q^n \end{pmatrix}, \quad a_n = \sqrt{n} \tag{2.37}$$

where p^n and q^n are the empirical distributions from simulating from p and q, and C is the set of all possible pairs of marginal distributions.

In order to carry out this strategy, we need to show that the optimal value of the dual objective function is Hadamard directionally differentiable tangent to C.

We have already shown that M_D is Gâteaux directionally differentiable in (2.34), so to ensure Hadamard directional differentiability, it is enough to show that M_D is locally Lipschitz continuous on C. But, for (p', q') in C, and $\bar{k} \in I(p, q)$:

$$\begin{aligned} M_D(p,q) - M_D(p',q') &= \min_k(\phi^k \cdot p + \psi^k \cdot q) - \min_n(\phi^k \cdot p' + \psi^k \cdot q') && (2.38) \\ &\le \phi^{\bar{k}} \cdot p + \psi^{\bar{k}} \cdot q - \phi^{\bar{k}} \cdot p' - \psi^{\bar{k}} \cdot q' && (2.39) \\ &\le \|\phi^{\bar{k}}\| \cdot \|p - p'\| + \|\psi^{\bar{k}}\| \cdot \|q - q'\| && (2.40) \\ &\le K_p \cdot \|p - p'\| + K_q \cdot \|q - q'\| && (2.41) \end{aligned}$$

where we have used the triangle inequality in the third step to get $\|p\| - \|p'\| \leq \|p - p'\|$, and $K_p = \max_k \|\varphi_k\|, K_q = \max_k \|\psi_k\|$. Switching the roles of (p, q) and (p', q') demonstrates that the maximum ES is Lipschitz on C, and thus Hadamard directionally differentiable, so the Delta method can be applied. We have therefore proved the limit result for the maximum ES:

Theorem 2. *Let C be the set of all (p, q) where p and q are probability vectors in $\mathbb{R}^M$ and $\mathbb{R}^N$. Suppose that p_n, q_n are the empirical measures from independent random samples from p and q. Then*

$$\sqrt{n}(M_D(p_n, q_n) - M_D(p, q)) \rightsquigarrow M'_D(p, q, Z_1, Z_2) \tag{2.42}$$

where:

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N(0, \Sigma) \tag{2.43}$$

with Σ as in (2.17).

Proof. This result follows from Theorem 1 and equation 2.34, using the fact that we have already shown that M_D is Hadamard differentiable tangent to C and that the derivative coincides with the Gateaux directional derivative when it exists. $\square$

Chapter 3

Numerical results and discussion

In this chapter, I will describe the method used in simulations, and describe the results. The full simulation code is available as an appendix. To begin, we look at a simple example where there is only one counterparty and both market factors Y and credit factors Z are standard normal random variables, and the loss function is their sum $L = Y + Z$. This provides an opportunity to see the limit theorem in a simple situation, where we can solve explicitly for the limit distribution and mean maximum ES. After that, a more financially relevant example of a portfolio with two counterparties will be examined and the empirical limiting distribution examined for various combinations of correlation, mean returns, variance, and default probability.

3.1 Simulation method

In order to simulate the limiting distribution linear programming problem described in Chapter 2, we use Matlab to write and run a script that calls the IBM CPLEX optimization studio's linear program solver. The linear program to be solved is

$$\max_{\pi,\Theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (L_{ij}\Theta_{ij} + 0 \cdot \pi_{ij}) \tag{3.1}$$

$$\sum_{j=1}^{n} \pi_{ij} = p_i, \; i \in \{1, \ldots, m\} \tag{3.2}$$

$$\sum_{i=1}^{m} \pi_{ij} = q_j, \; j \in \{1, \ldots, n\} \tag{3.3}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \Theta_{ij} = 1 \tag{3.4}$$

$$\Theta_{ij} \leq \frac{\pi_{ij}}{(1-\alpha)}, \; i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\} \tag{3.5}$$

$$\pi_{ij}, \Theta_{ij} \geq 0, \; i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\} \tag{3.6}$$

where p_i, q_j are the probabilities of the ith market scenario and the jth credit scenario. In our simulation, we take a Monte Carlo approach and simulate m market scenarios and n credit scenarios, so the probabilites are $p_i = \frac{1}{m} \; \forall i \in \{1, \ldots, m\}$, and $q_j = \frac{1}{n} \; \forall j \in \{1, \ldots, n\}$.

The method of simulating the losses is straightforward, we simulate two standard normal random variables Y and Z using one of two methods, either by calling Matlab's *randn* function or by simulating a random variable from a uniform $u \sim U(0, 1)$ distribution and taking the inverse normal CDF using Matlab's *norminv* function. As *cplexlp* solves for the minimum, we multiply the coefficients by -1 to get the inverse problem and then multiply the optimal value by -1 at the end to get the maximum.

The script is then run by a call that specifies the number of times each distribution is sampled (m and n), and the optimal value of the objective function is saved, along with the optimal values of the parameters for the primal and dual problems. The simulation is repeated N times, giving N different optimal values for the maximum ES. Some trial and error was needed to arrive at values of the parameters that offered a reasonable tradeoff between reproducible results and simulation time; this is discussed in more detail in the section describing the results.

Some considerations about the practicality of computation are worth mentioning at this juncture. There are two major limitations on the size of the simulation (that is, the values of m, n, and N). The problem has $2mn$ parameters in $mn + m + n + 1$ constraints, and therefore grows very quickly with larger values of m and n. There are also $2mnN$

values in the output of the program for both the primal and dual variables. The first consideration means that larger simulations take much longer to solve, since the size of the program grows like m^2 if m and n are of the same order. The second consideration relates to memory availability - the solution for all the programs for $m, n, N > 100$ can quickly rise into the multi-gigabyte range.

It is possible to ameliorate one problem slightly at the expense of worsening the others. Matlab allows the use of sparse matrices, which cut down on the amount of memory needed to store a matrix if its entries are mostly zeros. The tradeoff is that behind the scenes, Matlab doesn't reserve any memory for the matrix entries and every time a new entry is added the memory needs to be reassigned. This makes the program run more slowly for the benefit of using less memory (so long as the matrix is still *mostly* sparse, at any rate). Matlab can, generally, efficiently "sparsify" a matrix and combine sparse matrices, so if one knows in advance what it will look like one can then construct sub-matrices that do not use too much memory, sparsify them, and then concatenate them.

In our case, this works for the constraint matrices and using this method reduces computation time by a factor of five compared with adding non-zero values one by one to a pre-existing sparse matrix. The matrix of optimal values is impervious to this, however, and furthermore often does not contain enough zeros for it to be worth using sparse matrices. The result is that the limiting factor on simulation size was a combination of the length of time required (the script took around 18 hours to run for $m = 400,\ n = 400,\ N = 1000$) and the memory required for the solution.

An alternative method of solving the program above is to solve the dual problem, since the optimal values of the two problems will coincide. The dual of the maximum ES problem is

$$\min_{\phi,\psi,\rho,\beta} \sum_{i=1}^{m} \phi_i p_i \sum_{j=1}^{n} \psi_j q_j + \beta \tag{3.7}$$

$$(1-\alpha)(\phi_i - \psi_j) - \rho_{ij} \geq 0,\ i \in \{1,\dots,m\}, j \in \{1,\dots,n\} \tag{3.8}$$

$$\rho_{ij} + \beta \geq L_{ij},\ i \in \{1,\dots,m\}, j \in \{1,\dots,n\} \tag{3.9}$$

$$\rho_{ij} \geq 0,\ i \in \{1,\dots,m\}, j \in \{1,\dots,n\} \tag{3.10}$$

as described in Chapter 2. The dual problem, by comparison with the primal, has $mn + m + n + 1$ variables and $2mn$ constraints. We can, however, notice that the dual is a standard optimal transport problem with an added parameter β, and re-write it as

$$\min_{\phi,\psi} \sum_{i=1}^{m} \phi_i \mu_i \sum_{j=1}^{n} \psi_j \nu_j \tag{3.11}$$

$$\phi_i + \psi_j \geq C_i j^{\alpha,\beta} \tag{3.12}$$

where $C_{ij}^{\alpha,\beta} = (1-\alpha)\max(L_{ij} - \beta, 0)$. This gives a primal problem of the form

$$\max_{\pi} \sum_{i=1}^{m} \sum_{j=1}^{n} C_{i,j}^{\alpha,\beta} \pi_{ij} \sum_{j=1}^{n} \pi_{ij} = p_i,\ i \in \{1, \ldots, m\} \tag{3.13}$$

$$\sum_{i=1}^{m} \pi_{ij} = q_j,\ j \in \{1, \ldots, n\} \tag{3.14}$$

$$\pi_{ij} \geq 0 \tag{3.15}$$

The size of the problem is still of order m^2 (again assuming m and n are of the same order), but it is about half the size (assuming m and n are big enough, which here they always are), with mn variables and $m + n$ constraints.

There is no free lunch, however, and the challenge is finding the value of β that minimizes the optimal value of the primal problem. We have, effectively, gone from having to solve a larger problem once to solving a smaller problem multiple times. One approach is to perform a golden section search to try to find the best value of β. This runs into two challenges. The first is that the loss function is stochastic and therefore it's not easy to see when one is approaching the truly optimal value of β, so the sensitivity of the golden section search should be set to be not too restrictive, since later solutions of the same problem won't necessarily have the same optimal value for β.

The second challenge is that the solution will take much more time if the initial feasible point is far from the optimal solution, so some way of choosing an initial feasible point to minimize the solution time is required. This can be done by giving the solver the value of the previous optimal variables as the initial feasible point for the next iteration in the golden section search. Unfortunately, as will be discussed below, this particular problem is relatively impervious to this type of "hot starting".

In the end, while solving the the linear program was much faster (approximately 6-fold decrease in time for the solver to run) the time gained was consumed by the golden section search and the actual gains were not usually significant. Sometimes the solution was much faster and at othertimes it was slightly slower than the formulation presented in Chapter 2, depending on how many iterations the golden section search required but also, presumably, on how close to optimality the initial feasible point chosen was. In the end, the conclusion was that it did not offer consistent increases in speed. Nevertheless, if a consistent way of finding an initial feasible point for the solution in each iteration of the golden section search can be found, this approach still holds promise, although it is not the method of simulation used below.

3.2 Limit Distribution: Expected shortfall when loss is the sum of two standard normal random variables

We begin by considering the simple case of two normal random variables, $X \sim N(0,1)$ and $Y \sim N(0,1)$. Let the loss function be given by $L_{ij} = x_i + y_j$, where x_i and y_j are the ith and jth values drawn from X and Y [1]. This provides a straightforward example to both study the asymptotic distribution and ensure that the results of the code are as expected.

In this case, we can use the following fact:

$$ES(X+Y) \leq ES(X) + ES(Y) \tag{3.16}$$

since expected shortfall is sub-additive (see Chapter 1). This means that the maximum ES of the sum is equal to the sum of the individual ES of each factor when they are perfectly correlated, so we expect the worst-case ES to be equal to $ES(X) + ES(Y)$.

The ES of a standard normal is easily computed as

$$ES_\alpha(X) = \mu_x + \sigma_x \frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha} \tag{3.17}$$

Furthermore, we have that $(X+Y) \sim N(0,4)$ for perfectly correlated standard normal random variables. Evaluating with $\alpha = 0.9$ gives a value of 3.51 for the maximum ES.

[1] Note that the limit theorem from the previous chapter does not apply here as X and Y can both take on infinitely many values.

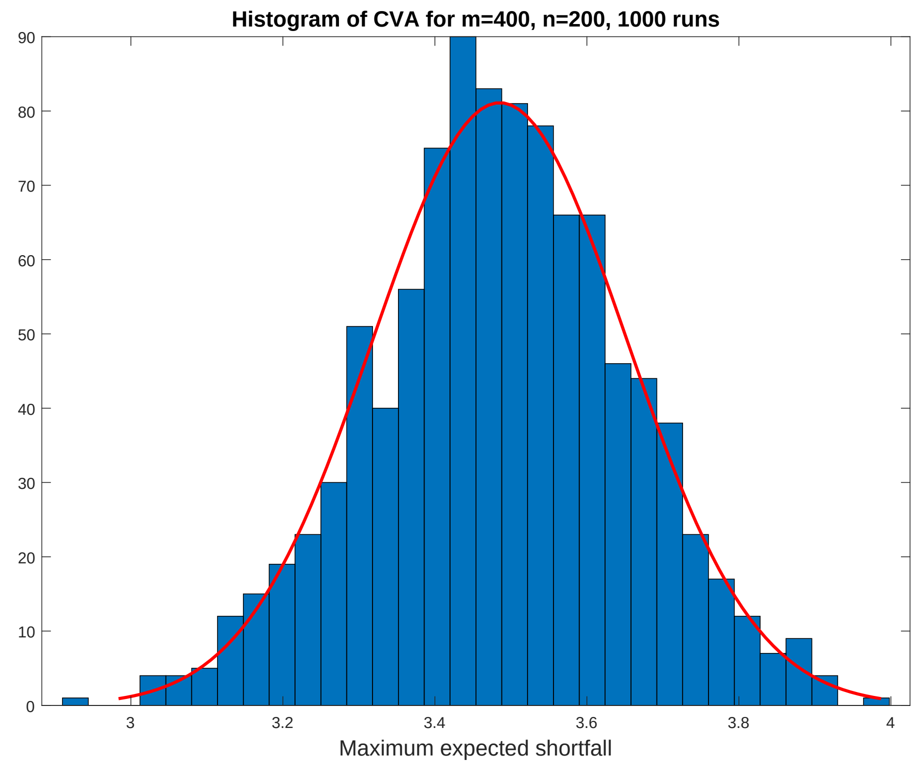

Figure 3.1: Histogram of the optimal value of the maximum ES problem when losses are given by $L = X + Y$. The red line indicates a fit to the normal distribution.

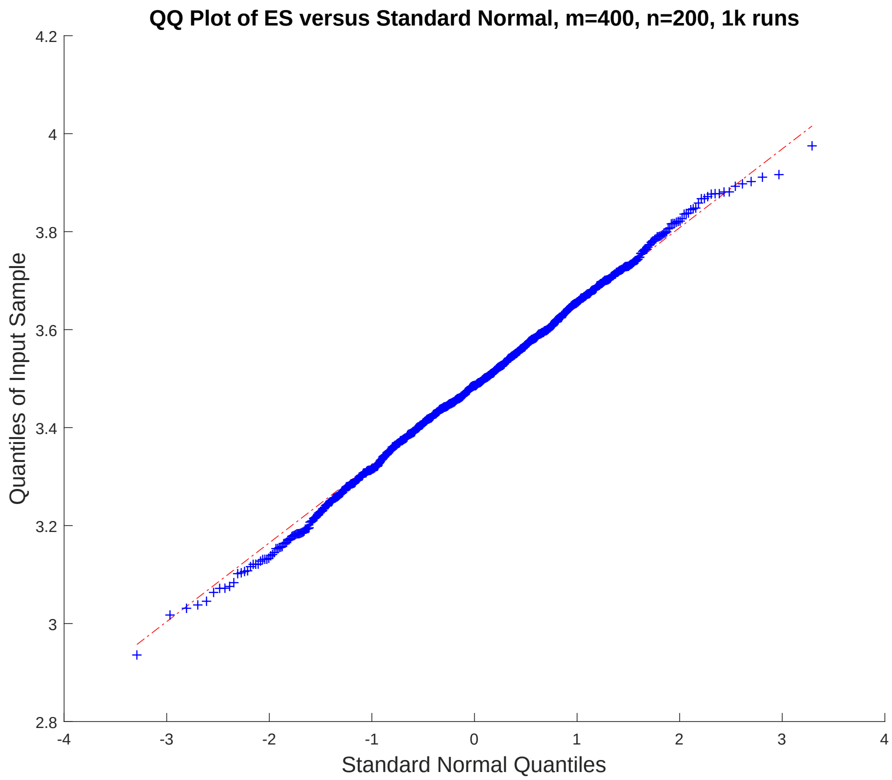

Figure 3.2: Quantiles of the optimal value of the maximum ES vs the quantiles of the normal distribution. The plot is roughly linear, indicating good agreement between the data and the normal distribution.

The corresponding worst-case ES simulation was run for 500 iterations with $\alpha = 0.9$ and the results for the optimal values of the linear program are plotted in Figure 3.1. The histogram is centred around the expected mean value of ~ 3.5 and a normal distribution fit is included for comparison. A simple Anderson-Darling test (Matlab's *adtest* function) offers $p = 0.92$ and fails to reject the hypothesis that the distribution is normal. A Q-Q plot of the results (Matlab's *qqplot*) vs the standard normal quantiles is shown in Figure 3.2. The linearity of the graph is good evidence that the distribution is indeed normal.

3.3 Limit distribution: Expected shortfall in the two counterparty portfolio

We now proceed to a more financially relevant example. We take as our starting point the case of a portfolio consisting of various positions with two counterparties, and proceed to describe the empirical limiting distribution of the maximum ES. The loss function is now as described in Chapter 2 for the Basel credit model.

The simulation takes a multivariate normal distribution for the market factors and a univariate normal for the credit factors. The market factors are characterized by their mean $\mu = [\mu_1, \mu_2]$ and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{3.18}$$

The parameters of the simulation are the correlation r, the average returns μ_1 and μ_2, the variances σ_1^2 and σ_2^2, and the default probabilities PD_1 and PD_2. The loading factor from equation (2.3) designated ρ is not included as a parameter and is rather set to a constant value of 0.2 to agree with standard literature values (see [12] for context and discussion) of ~ 0.45 for $\sqrt{\rho}$. Similarly, the value for α, the ES level, is fixed at 0.9 in order to strike a balance between getting the behaviour of the extreme values of VaR while still getting reasonable statistics.

The baseline values were taken to be $\mu_1 = \mu_2 = 100$, $\sigma_1 = \sigma_2 = 100$, and $PD_1 = PD_2 = 0.02$. The value of the correlation ρ in the covariance matrix is set to $\rho = 0.5$. These values were chosen not because they are particularly realistic but rather to exaggerate the salient features of the maximum distribution and give a qualitative characterization of the effects of changes of the parameters on the result. The histogram for this case is shown in Figure 3.3.

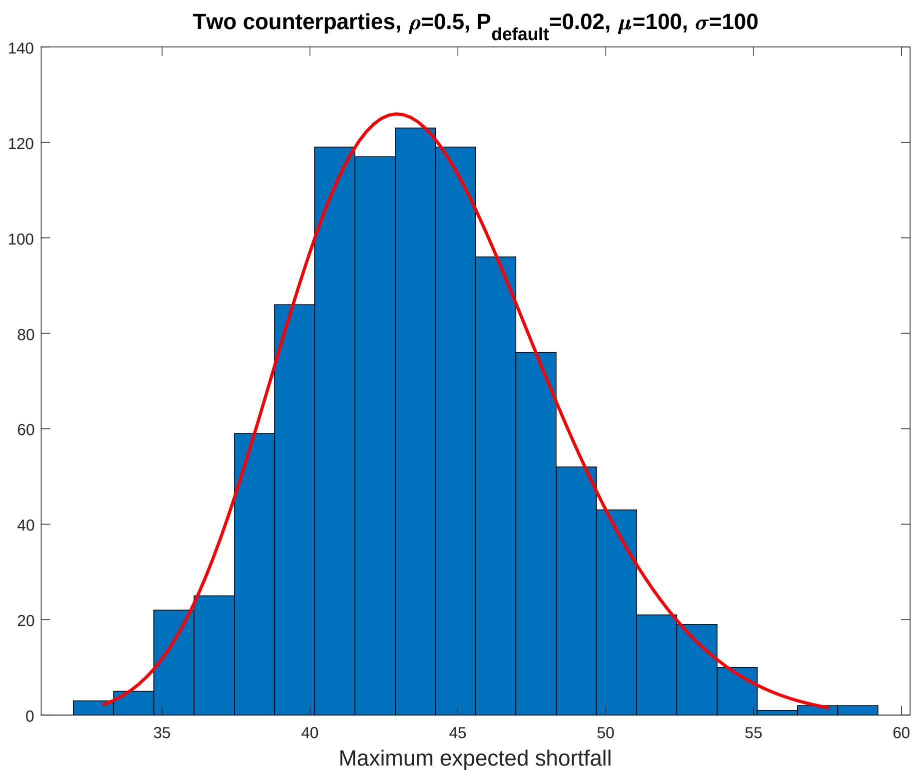

Figure 3.3: The histogram of maximum ES for the case of $\mu_1 = \mu_2 = 100$, $\sigma_1 = \sigma_2 = 100$, and $PD_1 = PD_2 = 0.02$. The red line indicates the fit to the GEV distribution.

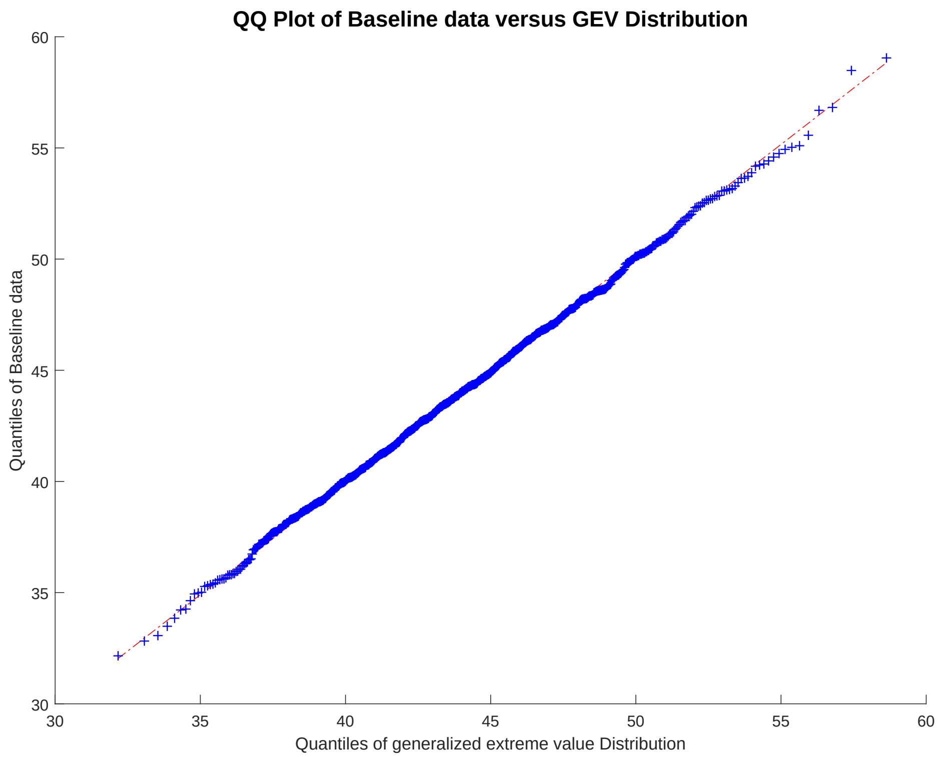

Figure 3.4: The Q-Q plot of maximum ES for the case of $\mu_1 = \mu_2 = 100$, $\sigma_1 = \sigma_2 = 100$, and $PD_1 = PD_2 = 0.02$ against the quantiles of the GEV distribution. The linearity of the graph suggests the distribution is well-described by the GEV distribution with the values indicated on the figure.

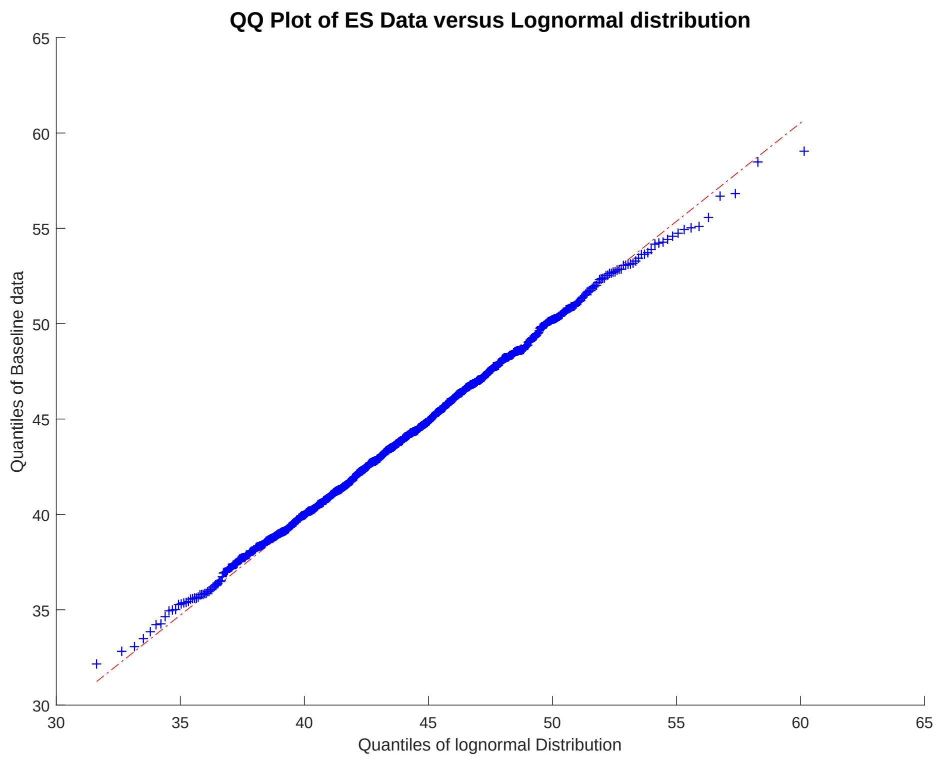

Figure 3.5: The Q-Q plot of maximum ES for the case of $\mu_1 = \mu_2 = 100$, $\sigma_1 = \sigma_2 = 100$, and $PD_1 = PD_2 = 0.02$ against the quantiles of the lognormal distribution. The linearity of the graph suggests the distribution is well-described by the LN distribution.

Evidently, the distribution shown here is not Gaussian. An Anderson-Darling test confirms this suspicion, returning $p = 0.0215$. Qualitatively, by inspection, the distribution is asymmetric about the mean, and the tails are too broad for the normal distribution to be a good candidate for a fit. By intuition, there are two distributions that one typically looks to with these properties that can be applied in situations in which one is interested in the empirical distribution: the lognormal (LN) distribution and the generalized extreme value (GEV) distribution. Both of these can be used to fit the data empirically.

The generalized extreme value fit to the data is shown in Figure 3.3. The Anderson-Darling test returns $p \sim 0.99$ for the GEV distribution and $p \sim 0.98$ for the LN distribution, indicating that both are viable empirical fits to the data. The shape parameter for the GEV distribution is -0.19, indicating a reverse Weibull type extreme value distribution.

The Q-Q plots for the results compared to the GEV and LN distributions are shown in Figure 3.4 and 3.5, and aside from some divergence at the extremes are both linear. This suggests that either one is a good choice for the fit to the data. For the remaining cases, we will only present the fit for the GEV distribution - intuitively, it is reasonable to expect that the solution to a maximization problem should conform to an extreme value distribution, even if a proof is lacking. Theorem 2, equation 2.41 and equation 2.34 from Chapter 2 show that when the optimal dual solution is degenerate, the limiting distribution is the maximum of a number of random variables, which suggests a GEV distribution is an appropriate way to model the result.

In addition, as stated previously, there are other (simpler) cases like linear regression where the limiting distribution of the solution of a linear program with stochastic coefficients corresponds to the GEV distribution [15]. Further, while great care must be taken when comparing distributions based on the p-value from the Anderson-Darling test, the p-value for the GEV distribution is consistently > 0.95 while that for the LN distribution gets worse for the other cases (though, admittedly, not worse than ~ 0.7).

The parameters are perturbed one at a time, and the results are shown in Figures 3.6, 3.8, and 3.10 with a fit to a GEV distribution. The GEV distribution fits the data well in all cases, resulting in shape parameters $\xi < 0$ that are similar and always in the Weibull-type regime. Values for maximum ES are shown on the Figures, along with the fit parameters for the GEV distribution. The Q-Q plots are shown as well, demonstrating that the GEV distribution describes the data well.

The plots show representative values for the parameters, rather than being an attempt to exhaustively categorize the limiting distributions and maximum ES. Simulations were done on smaller numbers of iterations (generally ~ 200) in order to get a sense of what effect the perturbation would have on the distribution. The changes that had a noteworthy

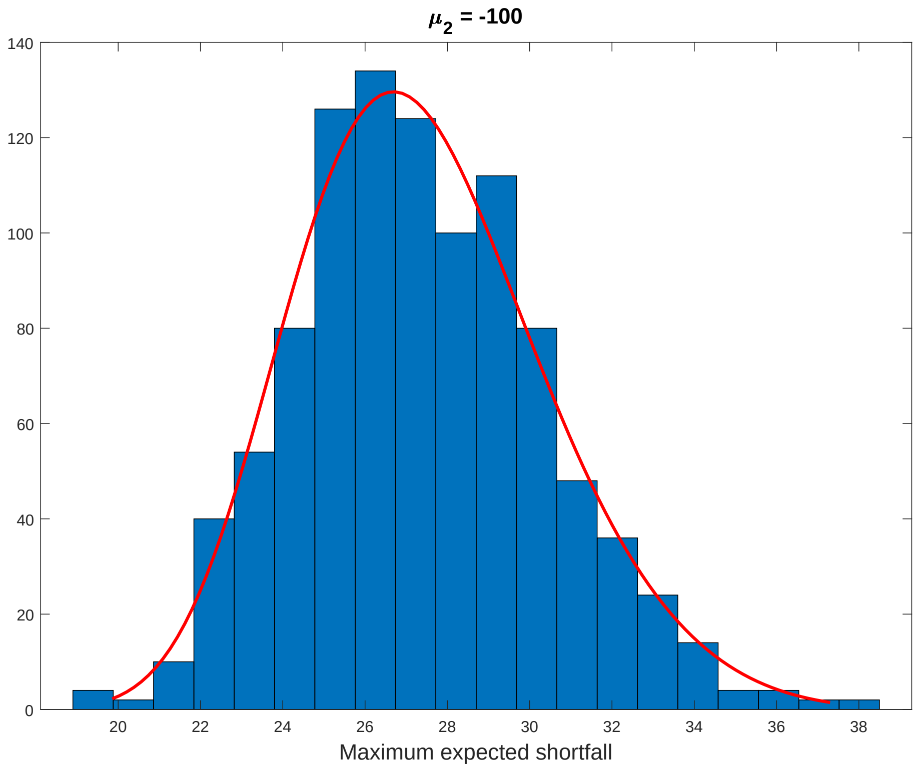

Figure 3.6: The histogram of maximum ES for the case of $\mu_1 = 100$, $\mu_2 = -100$. Other parameters are as in the base case. The red line indicates the fit to the GEV distribution.

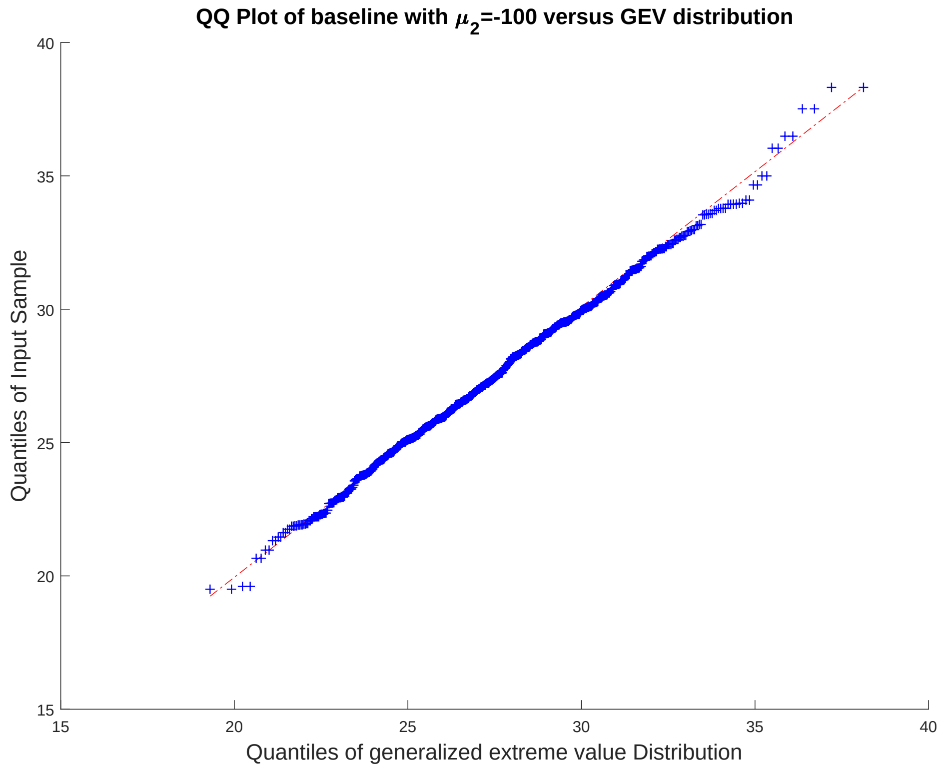

Figure 3.7: The Q-Q plot of maximum ES for the case of $\mu_1 = 100$, $\mu_2 = -100$ against the quantiles of the GEV distribution. The linearity of the graph suggests the distribution is well-described by the GEV distribution with the values indicated on the figure.

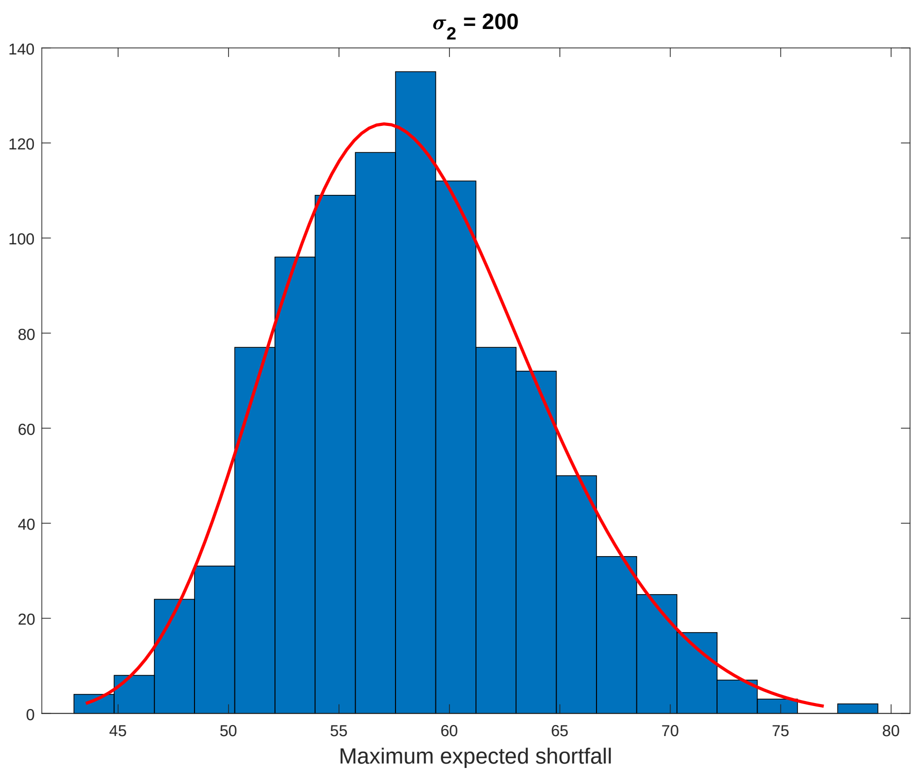

Figure 3.8: The histogram of maximum ES for the case of $\sigma_1 = 100$, $\sigma_2 = 200$. Other parameters are as in the base case. The red line indicates the fit to the GEV distribution.

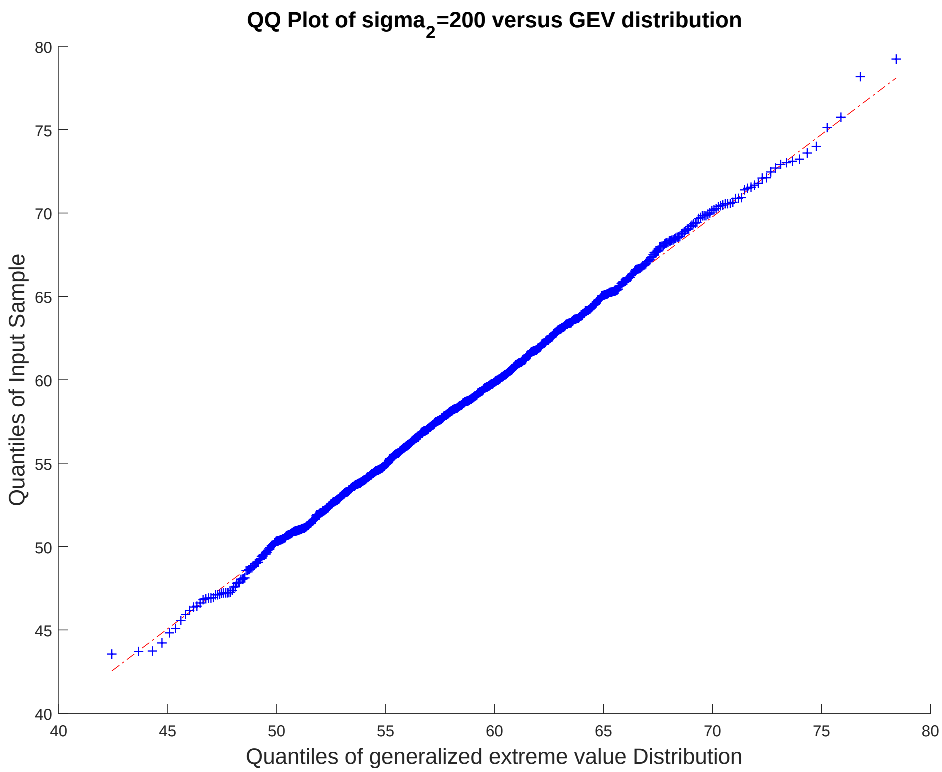

Figure 3.9: The Q-Q plot of maximum ES for the case of $\sigma_1 = 100$, $\sigma_2 = 200$ against the quantiles of the GEV distribution. The linearity of the graph suggests the distribution is well-described by the GEV distribution with the values indicated on the figure.

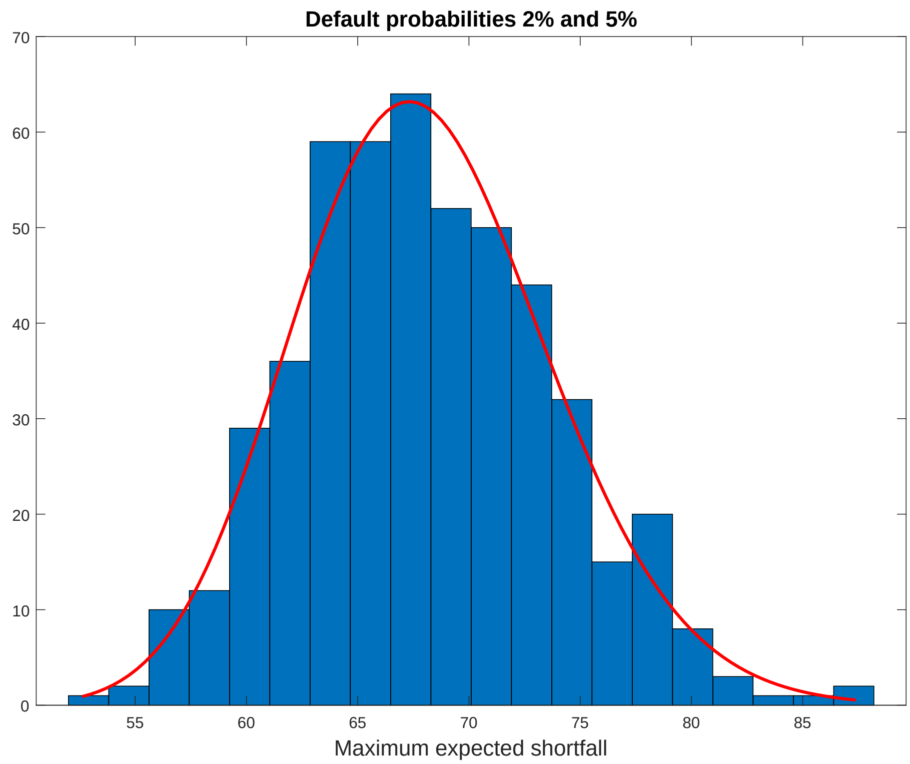

Figure 3.10: The histogram of maximum ES for the case of $PD_1 = 0.02$, $PD_2 = 0.05$. Other parameters are as in the base case. The red line indicates the fit to the GEV distribution.

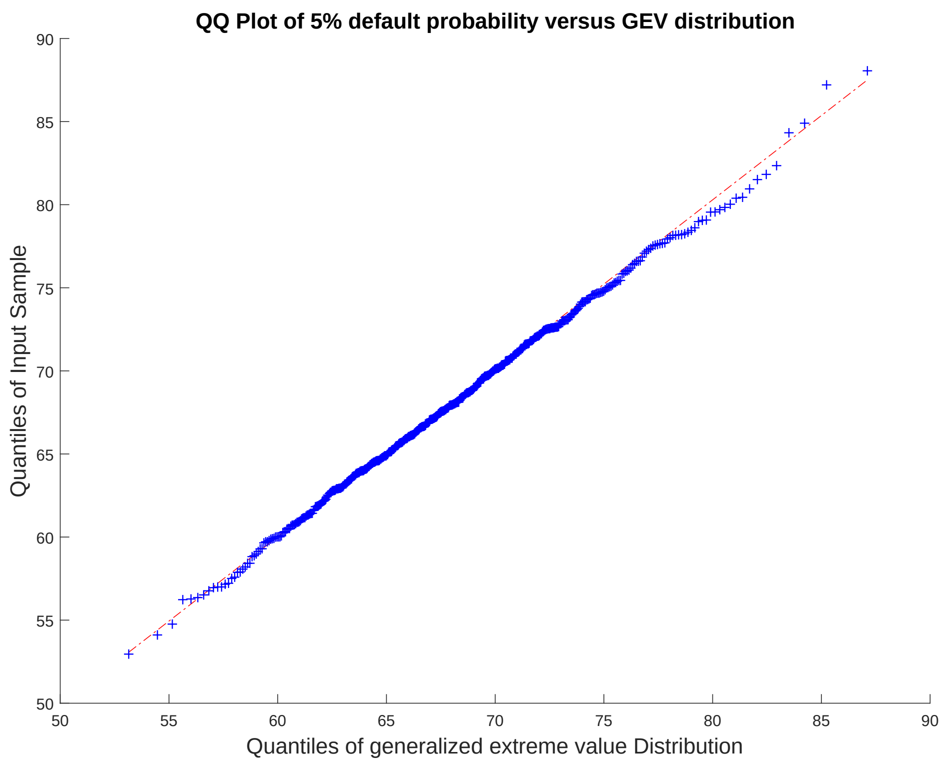

Figure 3.11: The Q-Q plot of maximum ES for the case of $PD_1 = 0.02$, $PD_2 = 0.05$ against the quantiles of the GEV distribution. The linearity of the graph suggests the distribution is well-described by the GEV distribution with the values indicated on the figure.

impact were then simulated 1000 times each in order to get better statistics for the fit. The exception is the case where $PD_2 = 0.05$, for which the simulation was done 1000 times but one of the data files became corrupted and only 500 values survived. As the result was not particularly strongly affected, in that the distribution is consistent with the others and the statistics are still clear, the figures with 500 values have been allowed to stand.

In particular, the correlation in the covariance matrix r did not have a significant effect on the output. Different combinations of the parameters being altered did not produce qualitative changes in the results; the limiting distributions were still well-described by the GEV distribution, and the mean maximum ES was consistent with the result of changing the parameters individually.

In order to characterize the stability of the result, we used the dual solution (output by the CPLEX solver as the lagrange multipliers) to determine the distance between the optimal solution points on the dual polytope. The hope was that the solution would look relatively stable, with a short distance between solutions. This hope was not realized due to the high degeneracy of the optimal dual solution; the average distance between optimal solutions was on the order of the norm of the individual optimal solutions. This does agree with the expectation that the result should be non-normal in the presence of degeneracy, however.

Chapter 4

Concluding remarks

This work has covered the problem of describing the limit distribution of expected shortfall for a loss function $L(Y, Z)$ when the marginal distributions of Y and Z are known but their joint distribution is unknown. The delta method, in combination with the Hadamard directional differentiability of the maximum ES tangent to the feasible set, allowed us to derive a limit theorem for the maximum expected shortfall. A simulation study was then presented considering the maximum ES for both the simple case of loss function that was the sum of standard normal random variables $L = Y + Z$ as well as a more financially relevant example of a portfolio of two counterparties with a loss function drawn from the Basel credit model.

The results here suggest that an extreme value distribution is the appropriate description for such cases. A full theoretical justification for this result remains unknown, but the empirical evidence, at least in the two-counterparty portfolio case, is reasonably compelling. This is not necessarily surprising, it has previously been shown that the generalized extreme value distribution appears as the limiting distribution for other linear programs [15]. The worst-case joint distribution problem that yields the maximum ES is an extreme value, and the generalized extreme value distribution is not an unreasonable guess for such a situation.

The cases considered here were relatively limited; two counterparties, with various combinations of values for the parameters. The simulation was mainly constrained by computing power, but larger portfolios, larger sample sizes, and different combinations of parameters could certainly be evaluated. Similarly, different distributions other than normal for the risk factors could be assessed to examine whether this changes the overall shape of the limiting distribution.

In this case, the market and credit risk factors were simulated from gaussian marginal distributions. Another extension of this work would be to follow the approach taken in Memartoluie *et al.* [18] and draw the credit and market scenarios from an actual portfolio or from historical data would offer insight into how well the results can be extended to real-world examples. If we know what the limiting distribution looks like in the case of actual portfolio data, the use of worst-case ES in practical risk-management applications could become a subject of interest.

References

[1] Alexandre Adam, Mohamed Houkari, and Jean-Paul Laurent. Spectral risk measures and portfolio selection. *Journal of Banking & Finance*, 32(9):1870–1882, 2008.

[2] Leif Andersen and Jakob Sidenius. Extensions to the gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk Volume*, 1(1):05, 2004.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[4] Basel Committee on Banking Supervision. Minimum capital requirements for market risk. Technical report, Bank for International Settlements, 2019. Available at www.bis.org.

[5] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17:477–501, 2013.

[6] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer, New York, second edition, 2006.

[7] Financial Crisis Inquiry Commission. *The financial crisis inquiry report: The final report of the National Commission on the causes of the financial and economic crisis in the United States including dissenting views*. Cosimo, Inc., 2011.

[8] E. del Barrio, E. Giné, and C. Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Annals of Probability*, 27(2):1009–1071, 1999.

[9] E. del Barrio, P. Gordaliza, and J.-M. Loubes. A central limit theorem for L_p transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference*, 8(4):817–849, 2019.

[10] J.C. Garcia-Cespedes, J.A. de Juan Herrero, D. Rosen, and D. Saunders. Effective modelling of wrong-way risk, CCR capital and alpha in Basel II. *Journal of Risk Model Validation*, 4(1):71–98, 2010.

[11] P. Glasserman and L. Yang. Bounding wrong-way risk in CVA calculation. *Mathematical Finance*, 28:268–305, 2018.

[12] Paul Glasserman and Jingyi Li. Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656, 2005.

[13] M. Klatt, C. Tameling, and A. Munk. Empirical regularized optimal transport: Statistical theory and applications. Available at arxiv.org, 2019.

[14] Marcel Klatt, Axel Munk, and Yoav Zemel. Limit laws for empirical optimal solutions in stochastic linear programs, 2020.

[15] Keith Knight. Limiting distributions of linear programming estimators. *Extremes*, 4(2):87–103, 2001.

[16] D.G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, Cham, fourth edition, 2016.

[17] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton University Press, Princeton, second edition, 2015.

[18] A. Memartoluie, D. Saunders, and T. Wirjanto. Wrong-way risk bounds in counterparty credit risk management. *Journal of Risk Management in Financial Institutions*, 10(2):150–163, 2017.

[19] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems*. Springer, New York, 1998.

[20] R.T. Rockefellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

[21] W. Römisch. Delta method, infinite dimensional. In S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic, and N.L. Johnson, editors, *Encyclopedia of Statistical Sciences*. John Wiley & Sons, 2006.

[22] D. Rosen and D. Saunders. Computing and stress testing counterparty credit risk capital. In E. Canabarro, editor, *Counterparty Credit Risk*, pages 245–292. RiskBooks, 2010.

[23] R.Y. Rubinstein and A. Shapiro. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. John Wiley & Sons, Chichester, 1993.

[24] L. Rüschendorf. *Mathematical Risk Analysis*. Springer, Berlin, 2013.

[25] Alexander Schied*. Risk measures and robust optimization problems. *Stochastic Models*, 22(4):753–831, 2006.

[26] A. Shapiro. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487, 1990.

[27] Richard L Smith. Nonregular regression. *Biometrika*, 81(1):173–183, 1994.

[28] M. Sommerfeld and A. Munk. Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B*, 80(1):219–238, 2018.

[29] C. Tameling, M. Sommerfeld, and A. Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. Available at arxiv.org, 2017.

[30] Oldrich Vasicek. Limiting loan loss probability distribution. *KMV corporation*, 1991.

[31] C. Villani. *Optimal Transport, Old and New*. Springer, Berlin, 2008.

APPENDICES

Appendix A

Code for normal loss function

```
function [X, fX, Aeq] = toy_model(m, n, alpha)

%     % Get the probabilities for the interval
%     x=linspace(-5, 5, 1000);
%     fn=zeros(1000,1);
%     for ii=2:999
%         fn(ii)=normcdf(x(ii+1))-normcdf(x(ii));
%     end
%     fn(1)=normcdf(-5);
%     fn(1000)=1-normcdf(5);

    % Generate independent uniform random variables
    U=rand(m, 1);
    V=rand(n,1);

    Y=norminv(U);
    Z=norminv(V);

    L=zeros(m*n,1);
    for ii=1:m
        for jj=1:n
            L((ii-1)*n+jj)=Y(ii) + Z(jj);
        end
```

```
end

p=1/m*ones(m,1);
q=1/n*ones(n,1);

Aeq = sparse(m+n+1, 2*m*n);
beq = [p;q;1];

% Equality constraint for sum(Theta) = 1
Aeq(m+n+1,:) = [zeros(1, m*n), ones(1, m*n)];

% Equality constraints, for sum_j pi_ij = p_i
for ii=1:m
    for jj=1:n
        Aeq(ii, (ii-1)*n+jj) = 1;
    end
end

% Equality constraints for sum_i pi_ij = q_j
for ii=m+1:m+n
    for jj=1:n:n*m
        Aeq(ii, jj + (ii - m - 1) ) = 1;
    end
end

% Inequality constraints, -1/(1-alpha)*Theta_ij + pi_ij <= 0
b = sparse(m*n, 1);

a = -1/(1 - alpha);

A = [a.*speye(m*n), speye(m*n)];

% upper and lower bounds
lb = zeros(2*m*n, 1);
ub = ones(2*m*n, 1); % not needed, included so linprog doesn't get confused

% coefficients for the pi's are 0, coefficients for the Theta's are
% -1*L
```

```
    f = [zeros(m*n, 1); -1.*L];

    % specify initial feasible point such that Theta_ij = pi_ij = p_i * q_j
    X_0=zeros(m,n);
    for ii=1:m
        for jj=1:n
            X_0(ii,jj)=p(ii).*q(jj);
        end
    end
    X0=[X_0; X_0];

    % Solve...?
%     [C, X] = linprog(f, A, b, Aeq, beq, lb, ub);
    [X, fX] = cplexlp(f,A,b,Aeq,beq,lb,ub, X0);
    fX = -fX;
```

Appendix B

Code for solving the two counterparty maximum expected shortfall

```
function [X, fX, exit, output, lambda] = two_cpty(mu1, mu2, sig1, sig2,...
 rho, pd1, pd2, m, n)

    %Start off with generating exposures for the two assets in the
    %portfolio

    %tic

    alpha = 0.9;
    Sigma = zeros(2,2);
    Sigma(1,1) = sig1^2;
    Sigma(2,2) = sig2^2;
    Sigma(1,2) = rho*sig1*sig2;
    Sigma(2,1) = Sigma(1,2);

    mu = [mu1; mu2];

    Y = mvnrnd(mu, Sigma, m); %matrix of asset values
```

```
%Initialize some variables to suitable values here

loading = 0.2; %to match literature value of ~0.45 for sqrt(loading)

Z = randn(n,1); %systematic credit factor

Lmn = zeros(m*n, 1);
%toc

for ii=1:m
    for jj=1:n
        Lmn((ii-1)*n+jj) = Y(ii, 1) * normcdf((norminv(pd1)...
             - sqrt(loading)*Z(jj))/sqrt(1-loading)) ...
            +Y(ii, 2)*normcdf((norminv(pd2)...
             - sqrt(loading)*Z(jj))/sqrt(1-loading));
    end
end

%Remainder solves the linear program as per toy_model

L=Lmn;

p=1/m*ones(m,1);
q=1/n*ones(n,1);

beq = [p;q;1];

% Equality constraints, for sum_j pi_ij = p_i
aa=1;
bb=1;
uu = zeros(2*m*n, 1);
vv = zeros(2*m*n, 1);
kk = ones(2*m*n, 1);
for ii=1:m
    for jj=1:n
        uu(aa) = ii;
        vv(aa) = (ii-1)*n+jj;
        aa = aa + 1;
```

```
    end
end

% Equality constraints for sum_i pi_ij = q_j
for ii=m+1:m+n
    for jj=1:n:n*m
        uu(aa) = ii;
        vv(aa) = jj + (ii - m - 1);
        aa = aa + 1;
    end
end

Aeq = sparse(uu, vv, kk, m+n+1, 2*m*n);
%toc

% Equality constraint for sum(Theta) = 1
Aeq(m+n+1,:) = sparse([zeros(1, m*n), ones(1, m*n)]);

% Inequality constraints, -1/(1-alpha)*Theta_ij + pi_ij <= 0
b = sparse(m*n, 1);

a = -1/(1 - alpha);

A = [a.*speye(m*n), speye(m*n)];

% upper and lower bounds
lb = zeros(2*m*n, 1);
ub = ones(2*m*n, 1); % not needed, included so linprog doesn't get confused

% coefficients for the pi's are 0, coefficients for the Theta's are
% -1*L
f = [zeros(m*n, 1); -1.*L];

% specify initial feasible point such that Theta_ij = pi_ij = p_i * q_j
X_0=zeros(m,n);
for ii=1:m
    for jj=1:n
        X_0(ii,jj)=p(ii).*q(jj);
```

```
        end
    end
    X0=[X_0; X_0];

    % Solve...?
%      [C, X] = linprog(f, A, b, Aeq, beq, lb, ub);
    [X, fX, exit, output, lambda] = cplexlp(f,A,b,Aeq,beq,lb,ub, X0);
    %toc
    fX = -fX;
```